

Component==3

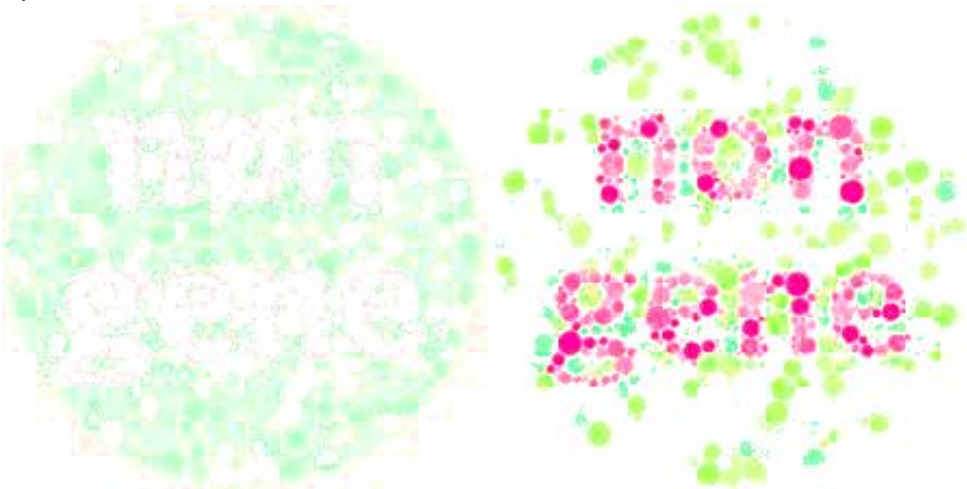
Question A:

After transformation:



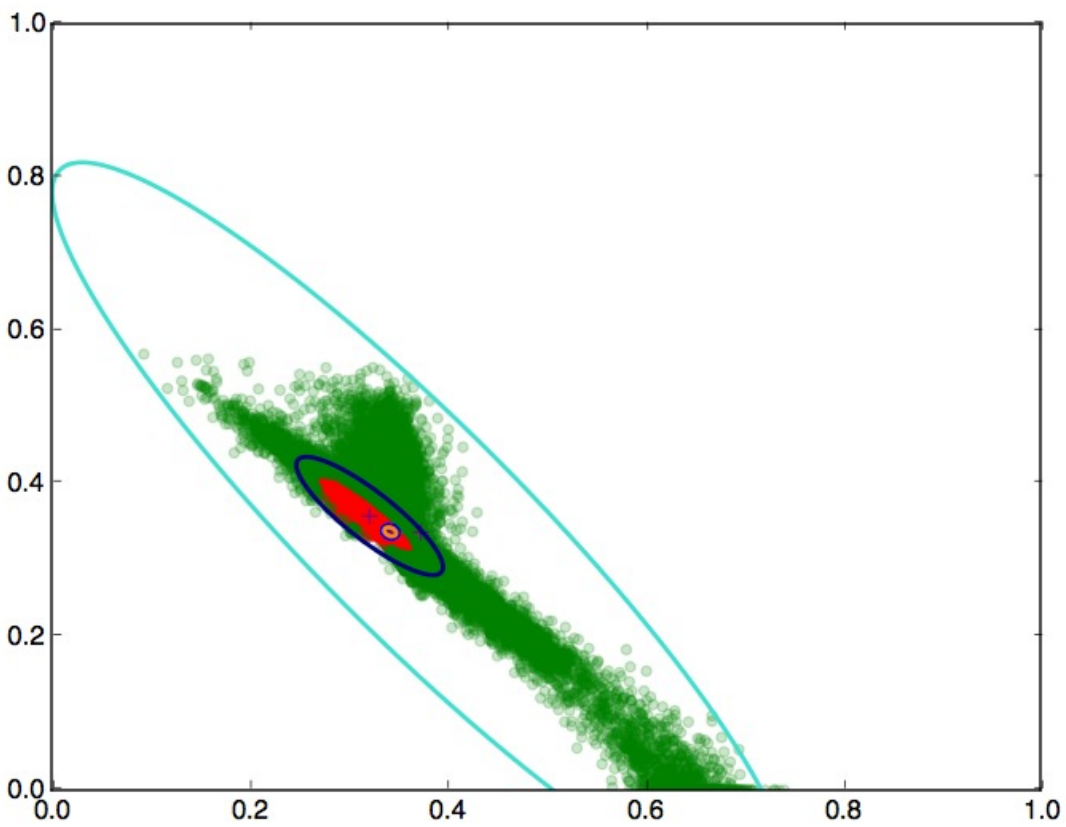
After transformation, some information is discarded. For example, green color is lightened. However, the word “nonGene” is still recognizable against the noises.

Question B



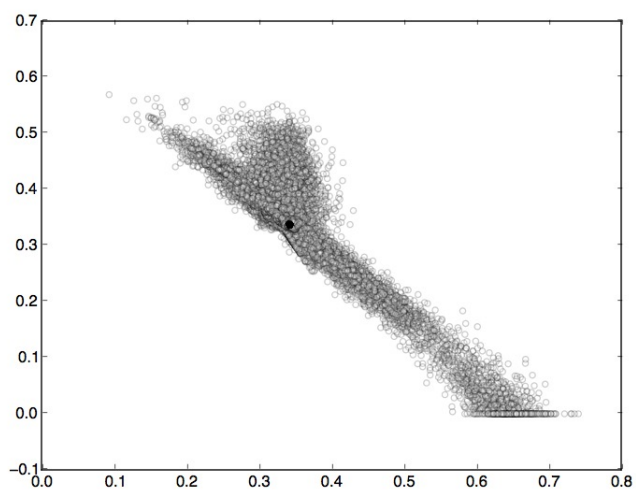
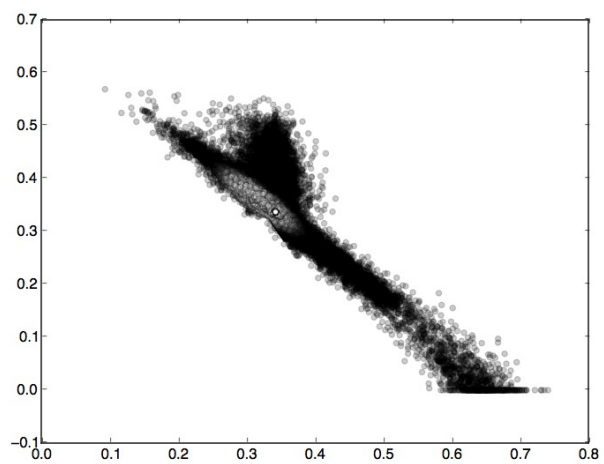
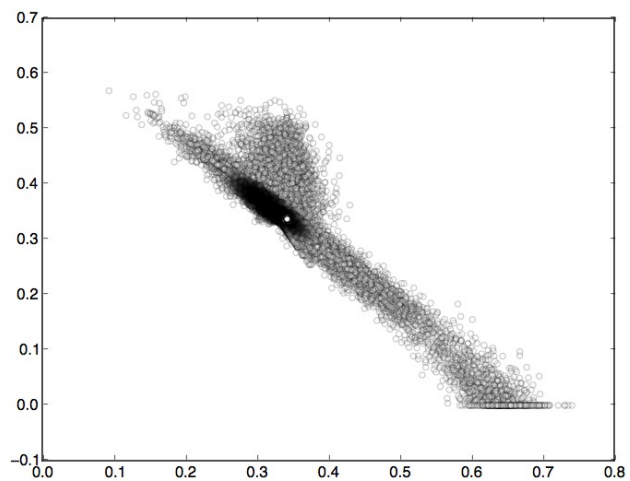


It appears that three cluster are well-separated.



In scatter plot, it looks like the the picture above. Where three clusters and its 3 sigma contour are shown, where the small inside circle is the third cluster.

li: three components with gray scaled color coding as probability is shown as:



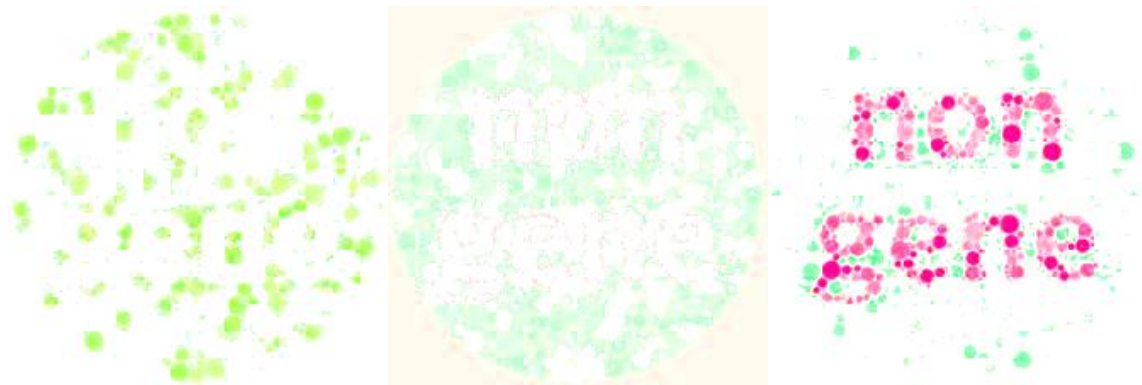
where the darkened area are the most likely area that a cluster resides. X,Y axis are the  $r$ , and  $g$  after transformation of question A.

iii

In general, all best Gaussian mixture model is selected by randomly choosing initial point, and select the highest likelihood. Three components are separated; the result looks good. Especially background is well separated based on the plot in B I. However, red layer and green noises are not separated. Some noises appeared in the graph. This is probably because after r,g transformation, the two color are very close in value. That result in the model unable to separate the difference.

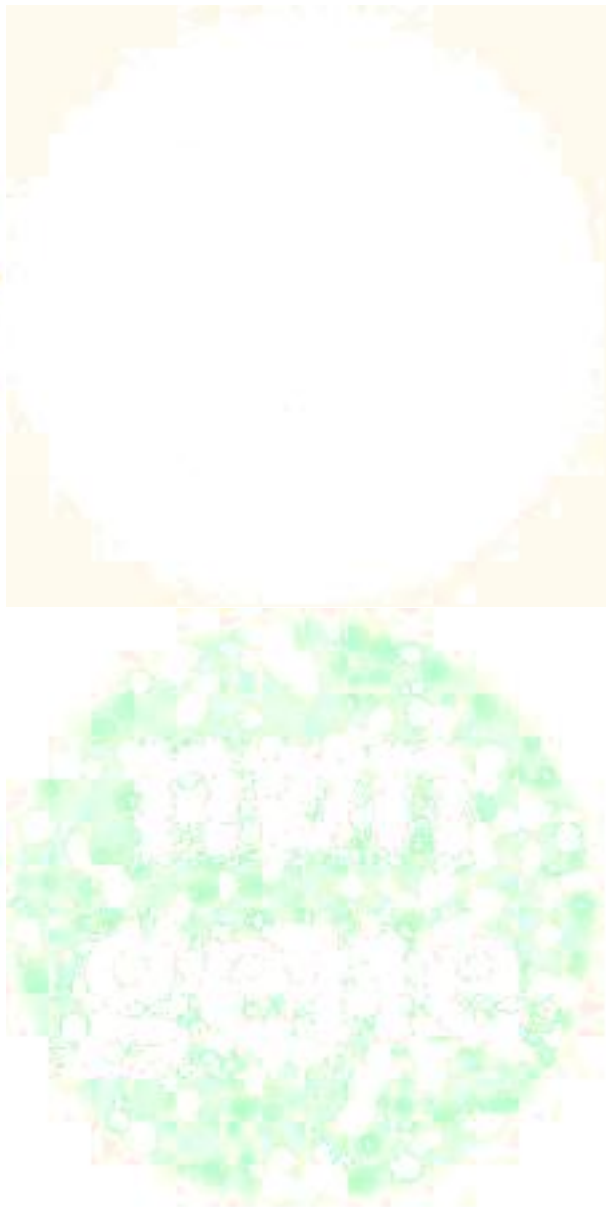
Another explanation is that the number of cluster is not enough, and in the original plot, there exists multiple type of noises. So that red and some type of noises have to be grouped together because cluster number limitation.

Initialization is also very important for this algorithm. Because of limited size of the page I will show only one random initialization with cluster size==3. Notice the result is not best selected by choosing highest likelihood among multiple models:

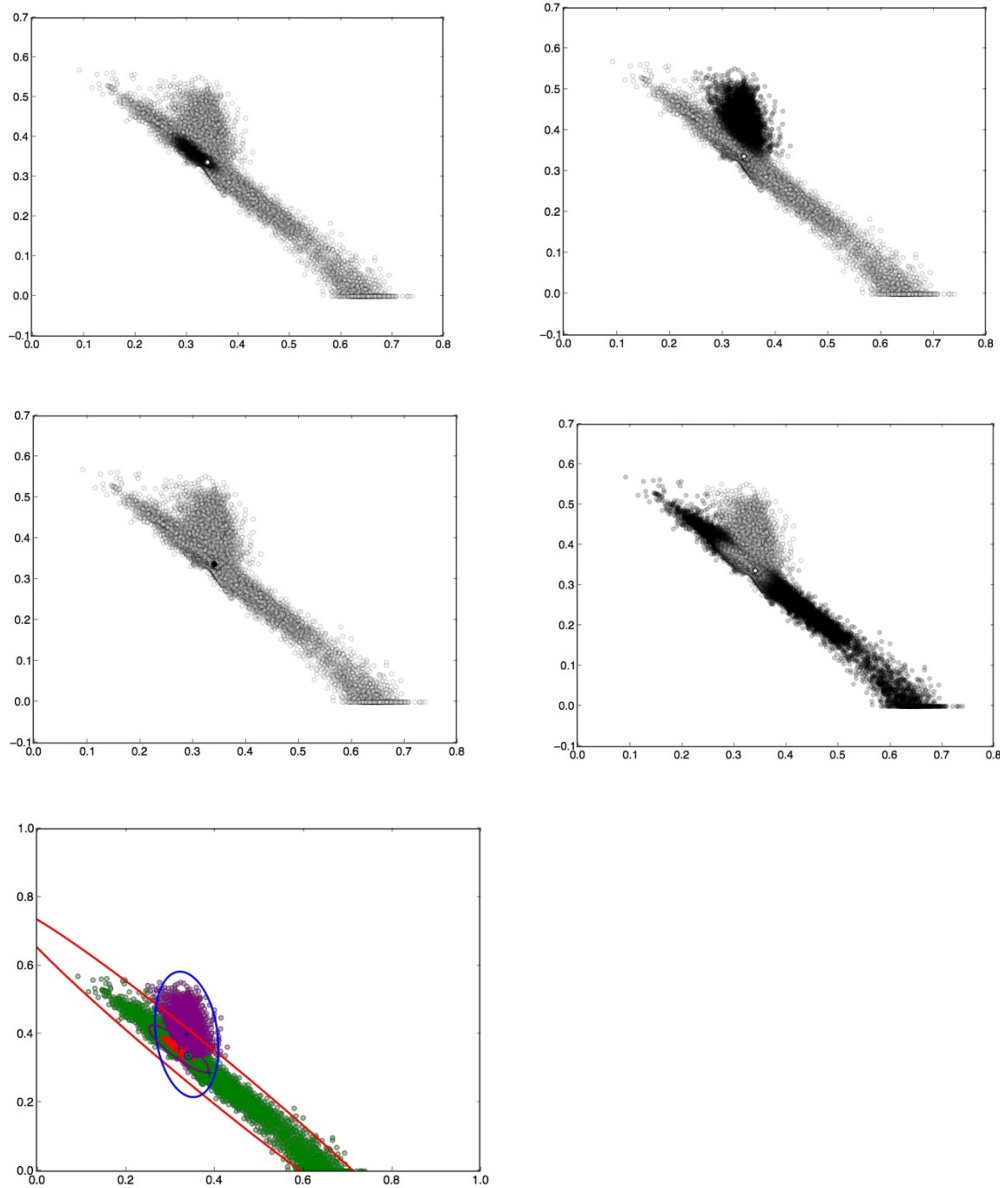


Replicate result with components ==4

The separation looks like:



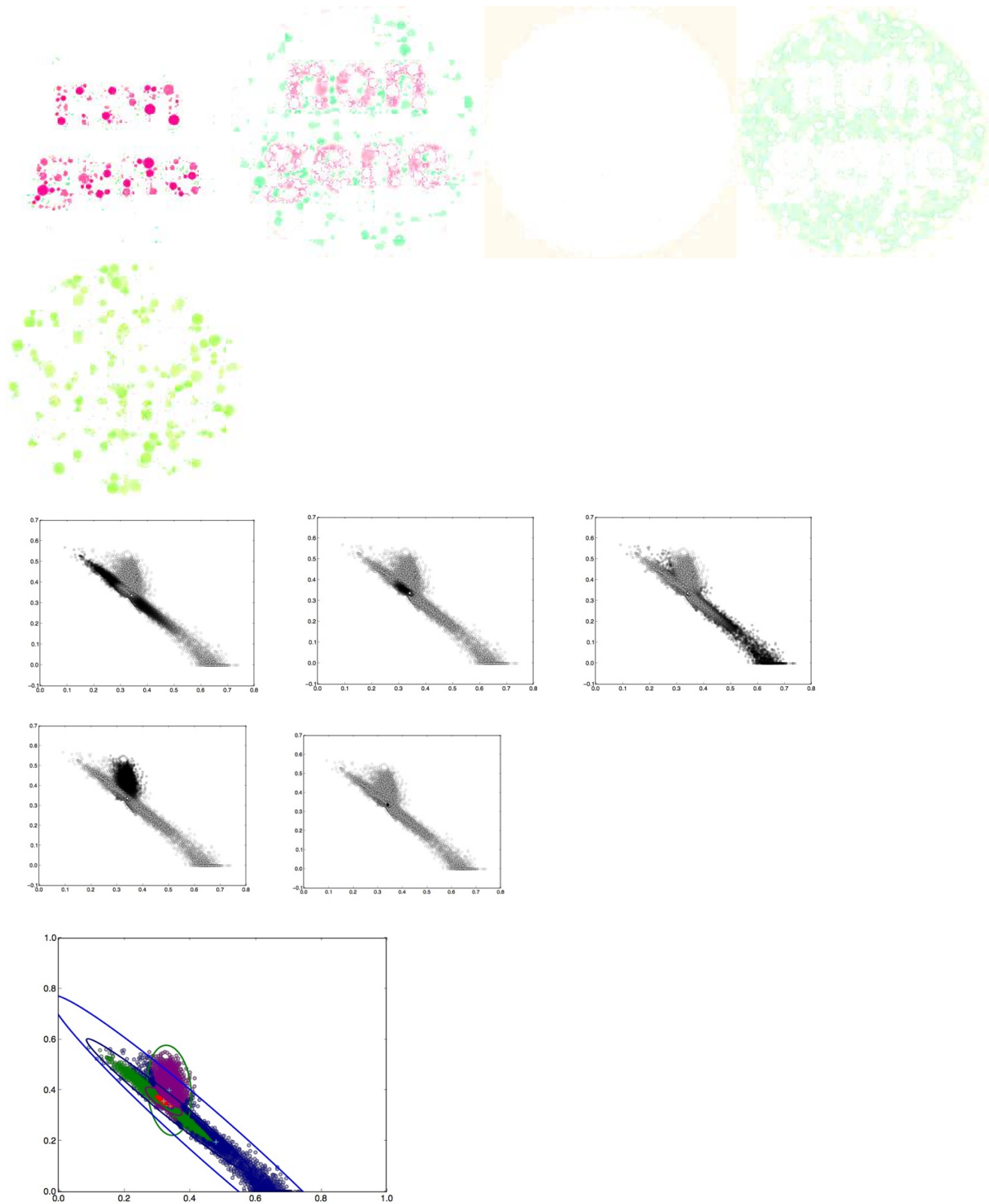
The clustering result probability distribution and contour looks like:



#### Comments:

The result looks slightly better compare to component==3 because blue layer is removed from signal. Again the initial condition matter, the plot shown above is selected from the highest likelihood. With other condition the separation is worse than this one. Because of limited printing resource, other 19 modes' result will not be shown.

Now cluster with component==5:



Comments:

Clearly, cluster ==5 separated the signal in two two channels. The result is no better than 4-component model.

In general, the best result is component==4. With broader coverage of initial points, it is possible to separate a better result